DIAGONAL AND LOW-RANK MATRIX DECOMPOSITIONS, CORRELATION MATRICES, AND ELLIPSOID FITTING*

J. SAUNDERSON†, V. CHANDRASEKARAN†, P. A. PARRILO†, AND A. S. WILLSKY†

Abstract. In this paper we establish links between, and new results for, three problems that are not usually considered together. The first is a matrix decomposition problem that arises in areas such as statistical modeling and signal processing: given a matrix X formed as the sum of an unknown diagonal matrix and an unknown low rank positive semidefinite matrix, decompose X into these constituents. The second problem we consider is to determine the facial structure of the set of correlation matrices, a convex set also known as the elliptope. This convex body, and particularly its facial structure, plays a role in applications from combinatorial optimization to mathematical finance. The third problem is a basic geometric question: given points $v_1, v_2, \ldots, v_n \in \mathbb{R}^k$ (where $n > k$) determine whether there is a centered ellipsoid passing *exactly* through all of the points.

We show that in a precise sense these three problems are equivalent. Furthermore we establish a simple sufficient condition on a subspace $\mathcal{U}$ that ensures any positive semidefinite matrix L with column space $\mathcal{U}$ can be recovered from $D+L$ for any diagonal matrix D using a convex optimization-based heuristic known as minimum trace factor analysis. This result leads to a new understanding of the structure of rank-deficient correlation matrices and a simple condition on a set of points that ensures there is a centered ellipsoid passing through them.

Key words. Elliptope, minimum trace factor analysis, Frisch scheme, semidefinite programming, subspace coherence

AMS subject classifications. 90C22, 52A20, 62H25, 93B30

1. Introduction.

Decomposing a matrix as a sum of matrices with simple structure is a fundamental operation with numerous applications. A matrix decomposition may provide computational benefits, such as allowing the efficient solution of the associated linear system in the square case. Furthermore, if the matrix arises from measurements of a physical process (such as a sample covariance matrix), decomposing that matrix can provide valuable insight about the structure of the physical process.

Among the most basic and well-studied additive matrix decompositions is the decomposition of a matrix as the sum of a diagonal matrix and a low-rank matrix. This decomposition problem arises in the factor analysis model in statistics, which has been studied extensively since Spearman's original work of 1904 [29]. The same decomposition problem is known as the Frisch scheme in the system identification literature [17]. For concreteness, in Section 1.1 we briefly discuss a stylized version of a problem in signal processing that under various assumptions can be modeled as a (block) diagonal and low-rank decomposition problem.

Much of the literature on diagonal and low-rank matrix decompositions is in one of two veins. An early approach [1] that has seen recent renewed interest [11] is an algebraic one, where the principal aim is to give a characterization of the vanishing ideal of the set of symmetric $n \times n$ matrices that decompose as the sum of a diagonal matrix and a rank k matrix. Such a characterization has only been obtained for the

*This research was funded in part by Shell International Exploration and Production, Inc. under P.O. 450004440, and in part by the Air Force Office of Scientific Research under grant #FA9550-11-1-0305. A preliminary version of parts of this work appeared in the Master's thesis of the first-named author [24].

†Laboratory for Information and Decision Systems, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 (jamess@mit.edu, venkatc@mit.edu, parrilo@mit.edu, willsky@mit.edu).

border cases $k = 1$, $k = n - 1$ (due to Kalman [17]), and the recently resolved $k = 2$ case (due to Brouwer and Draisma [3] following a conjecture by Drton et al. [11]). This approach does not (yet) offer scalable algorithms for performing decompositions, rendering it unsuitable for many applications including those in high-dimensional statistics, optics [12], and signal processing [24]. The other main approach to factor analysis is via heuristic local optimization techniques, often based on the expectation maximization (EM) algorithm [9]. This approach, while computationally tractable, typically offers no provable performance guarantees.

A third way is offered by convex optimization-based methods for diagonal and low-rank decompositions such as *minimum trace factor analysis* (MTFA), the idea and initial analysis of which dates at least to Ledermann's 1940 work [21]. MTFA is computationally tractable, being based on a semidefinite program (see Section 2), and yet offers the possibility of provable performance guarantees. In this paper we provide a new analysis of MTFA that is particularly suitable for high-dimensional problems.

Semidefinite programming duality theory provides a link between this matrix decomposition heuristic and the facial structure of the set of *correlation matrices*—positive semidefinite matrices with unit diagonal—also known as the *elliptope* [19]. This set is one of the simplest of spectrahedra—affine sections of the positive semidefinite cone. Spectrahedra are of particular interest for two reasons. First, spectrahedra are a rich class of convex sets that have many nice properties (such as being facially exposed). Second, there are well-developed algorithms, efficient both in theory and in practice, for optimizing linear functionals over spectrahedra. These optimization problems are known as semidefinite programs [30].

The elliptope arises in semidefinite programming-based relaxations of problems in areas such as combinatorial optimization (e.g. the MAX-CUT problem [14]) and statistical mechanics (e.g. the k-vector spin glass problem [2]). In addition, the problem of projecting onto the set of (possibly low-rank) correlation matrices has enjoyed considerable interest in mathematical finance and numerical analysis in recent years [16]. In each of these applications the structure of the set of low-rank correlation matrices, i.e. the facial structure of this convex body, plays an important role.

Understanding the faces of the elliptope turns out to be related to the following *ellipsoid fitting problem*: given n points in $\mathbb{R}^k$ (with $n > k$), under what conditions on the points is there an ellipsoid centered at the origin that passes *exactly* through these points? While there is considerable literature on many ellipsoid-related problems, we are not aware of any previous systematic investigation of this particular problem.

1.1. Illustrative application: direction of arrival estimation. Direction of arrival estimation is a classical problem in signal processing where (block) diagonal and low-rank decomposition problems arise naturally. In this section we briefly discuss some stylized models of the direction of arrival estimation problem that can be reduced to matrix decomposition problems of the type considered in this paper.

Suppose we have n sensors at locations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \in \mathbb{R}^2$ that are passively 'listening' for waves (electromagnetic or acoustic) at a known frequency from $r \ll n$ sources in the far field (so that the waves are approximately plane waves when they reach the sensors). The aim is to estimate the number of sources r and their directions of arrival $\theta = (\theta_1, \theta_2, \ldots, \theta_r)$ given sensor measurements and knowledge of the sensor locations (see Figure 1.1).

A standard mathematical model for this problem (see [18] for a derivation) is to model the vector of sensor measurements $z(t) \in \mathbb{C}^n$ at time t as

$$z(t) = A(\theta)s(t) + n(t) \tag{1.1}$$

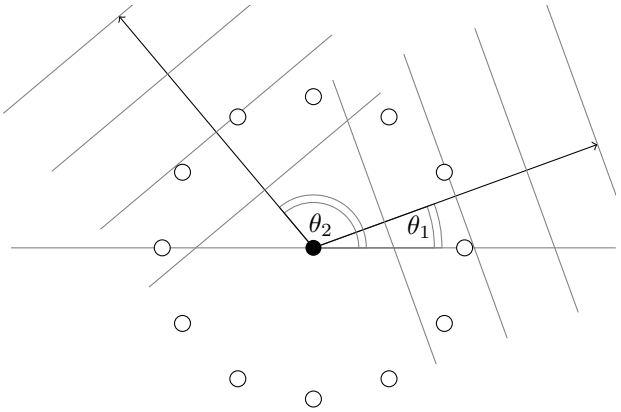


Fig. 1.1: Plane waves from directions θ_1 and θ_2 arriving at an array of sensors equally spaced on a circle (a uniform circular array).

where $s(t) \in \mathbb{C}^r$ is the vector of baseband signal waveforms from the sources, $n(t) \in \mathbb{C}^n$ is the vector of sensor measurement noise, and $A(\theta)$ is the $n \times r$ matrix with complex entries $[A(\theta)]_{ij} = e^{-k\sqrt{-1}(x_i \cos(\theta_j) + y_i \sin(\theta_j))}$, with k a positive constant related to the frequency of the waves being sensed.

The column space of $A(\theta)$ contains all the information about the directions of arrival θ. As such, subspace-based approaches to direction of arrival estimation aim to estimate the column space of $A(\theta)$ (from which a number of standard techniques can be employed to estimate θ).

Typically $s(t)$ and $n(t)$ are modeled as zero-mean stationary white Gaussian processes with covariances $\mathbb{E}[s(t)s(t)^H] = P$ and $\mathbb{E}[n(t)n(t)^H] = Q$ respectively (where A^H denotes the Hermitian transpose of A and $\mathbb{E}[\cdot]$ the expectation). In the simplest setting, $s(t)$ and $n(t)$ are assumed to be uncorrelated so that the covariance of the sensor measurements at any time is

$$\Sigma = A(\theta)PA(\theta)^H + Q.$$

The first term is Hermitian positive semidefinite with rank r, i.e. the number of sources. Under the assumption that spatially well-separated sensors (such as in a sensor network) have uncorrelated measurement noise Q is diagonal. In this case the covariance Σ of the sensor measurements decomposes as a sum of a positive semidefinite matrix of rank $r \ll n$ and a diagonal matrix. Given an approximation of Σ (e.g. a sample covariance) approximately performing this diagonal and low-rank matrix decomposition allows the estimation of the column space of $A(\theta)$ and in turn the directions of arrival.

A variation on this problem occurs if there are multiple sensors at each location, sensing, for example, waves at different frequencies. Again under the assumption that well-separated sensors have uncorrelated measurement noise, and sensors at the same location have correlated measurement noise, the sensor noise covariance matrix Q would be *block-diagonal*. As such the covariance of all of the sensor measurements would decompose as the sum of a low-rank matrix (with rank equal to the total number of sources over all measured frequencies) and a block-diagonal matrix.

A block-diagonal and low-rank decomposition problem also arises if the second-

order statistics of the noise have certain *symmetries*. This might occur in cases where the sensors themselves are arranged in a symmetric way (such as in the uniform circular array shown in Figure 1.1). In this case there is a unitary matrix T (depending only on the symmetry group of the array) such that TQT^H is *block-diagonal* [25]. Then the covariance of the sensor measurements, when written in coordinates with respect to T, is

$$T\Sigma T^H = TA(\theta)PA(\theta)^H T^H + TQT^H$$

which has a decomposition as the sum of a block diagonal matrix and a rank r Hermitian positive semidefinite matrix (as conjugation by T does not change the rank of this term).

Note that the matrix decomposition problems discussed in this section involve Hermitian matrices with complex entries, rather than the symmetric matrices with real entries considered elsewhere in this paper. It is straightforward to generalize the main problems and results throughout the paper to the complex setting.

1.2. Contributions.

Relating MTFA, correlation matrices, and ellipsoid fitting. We introduce and make explicit the links between the analysis of MTFA, the facial structure of the elliptope, and the ellipsoid fitting problem, showing that these problems are, in a precise sense, equivalent (see Proposition 3.1). As such, we relate a basic problem in statistical modeling (tractable diagonal and low-rank matrix decompositions), a basic problem in convex algebraic geometry (understanding the facial structure of perhaps the simplest of spectrahedra), and a basic geometric problem.

A sufficient condition for the three problems. The main result of the paper is to establish a new, simple, sufficient condition on a subspace $\mathcal{U}$ of $\mathbb{R}^n$ that ensures that MTFA correctly decomposes matrices of the form $D^\star + L^\star$ where $\mathcal{U}$ is the column space of $L^\star$. The condition is stated in terms of a measure of *coherence* of a subspace (made precise in Definition 4.1). Informally, the coherence of a subspace is a real number between zero and one that measures how close the subspace is to containing any of the elementary unit vectors. This result can be translated into new results for the other two problems under consideration based on the relationship between the analysis of MTFA, the faces of the elliptope, and ellipsoid fitting.

Block-diagonal and low-rank decompositions. In Section 5 we turn our attention to the *block*-diagonal and low-rank decomposition problem, showing how our results generalize to that setting. Our arguments combine our results for the diagonal and low-rank decomposition case with an understanding of the symmetries of the block-diagonal and low-rank decomposition problem.

1.3. Outline. The remainder of the paper is organized as follows. We describe notation, give some background on semidefinite programming, and provide precise problem statements in Section 2. In Section 3 we present our first contribution by establishing relationships between the success of MTFA, the faces of the elliptope, and ellipsoid fitting. We then illustrate these connections by noting the equivalence of a known result about the faces of the elliptope, and a known result about MTFA, and translating these into the context of ellipsoid fitting. Section 4 is focused on establishing and interpreting our main result: a sufficient condition for the three problems based on a coherence inequality. Finally in Section 5 we generalize our results to the analogous tractable block-diagonal and low-rank decomposition problem.

2. Background and problem statements.

2.1. Notation. If $x, y \in \mathbb{R}^n$ we denote by $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ the standard Euclidean inner product and by $\|x\|_2 = \langle x, x\rangle^{1/2}$ the corresponding Euclidean norm. We write $x \geq 0$ and $x > 0$ to indicate that x is entry-wise non-negative and strictly positive, respectively. Correspondingly, if $X, Y \in \mathcal{S}^n$, the set of $n \times n$ symmetric matrices, then we denote by $\langle X, Y\rangle = \mathrm{tr}(XY)$ the trace inner product and by $\|X\|_F = \langle X, X\rangle^{1/2}$ the Frobenius norm. We write $X \succeq 0$ and $X \succ 0$ to indicate that X is positive semidefinite and strictly positive definite, respectively. We write $\mathcal{S}^n_+$ for the cone of $n \times n$ positive semidefinite matrices.

The column space of a matrix X is denoted $\mathcal{R}(X)$ and the nullspace is denoted $\mathcal{N}(X)$. If X is an $n \times n$ matrix then $\mathrm{diag}(X) \in \mathbb{R}^n$ is the diagonal of X. If $x \in \mathbb{R}^n$ then $\mathrm{diag}^*(x) \in \mathcal{S}^n$ is the diagonal matrix with $[\mathrm{diag}^*(x)]_{ii} = x_i$ for $i = 1, 2, \ldots, n$. If $\mathcal{U}$ is a subspace of $\mathbb{R}^n$ then $P_{\mathcal{U}} : \mathbb{R}^n \to \mathbb{R}^n$ denotes the orthogonal projector onto $\mathcal{U}$, that is the self-adjoint linear map such that $\mathcal{R}(P_{\mathcal{U}}) = \mathcal{U}$, $P_{\mathcal{U}}^2 = P_{\mathcal{U}}$ and $\mathrm{tr}(P_{\mathcal{U}}) = \dim(\mathcal{U})$.

We use the notation e_i for the vector with a one in the ith position and zeros elsewhere and the notation $\mathbf{1}$ to denote the vector all entries of which are one. We use the shorthand $[n]$ for the set $\{1, 2, \ldots, n\}$. The set of $n \times n$ correlation matrices, i.e. positive semidefinite matrices with unit diagonal, is denoted $\mathcal{E}_n$. For brevity we typically refer to $\mathcal{E}_n$ as the elliptope, and the elements of $\mathcal{E}_n$ as correlation matrices.

2.2. Semidefinite programming. The term semidefinite programming [30] refers to convex optimization problems of the form

$$\underset{X}{\text{minimize}}\ \langle C, X\rangle \quad \text{subject to} \quad \begin{cases} \mathcal{A}(X) = b \\ X \succeq 0 \end{cases} \tag{2.1}$$

where X and C are $n \times n$ symmetric matrices, $b \in \mathbb{R}^m$, and $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$ is a linear map. The dual semidefinite program is

$$\underset{y,S}{\text{maximize}}\ \langle b, y\rangle \quad \text{subject to} \quad \begin{cases} C - \mathcal{A}^*(y) = S \\ S \succeq 0 \end{cases} \tag{2.2}$$

where $\mathcal{A}^* : \mathbb{R}^m \to \mathcal{S}^n$ is the adjoint of $\mathcal{A}$.

General semidefinite programs can be solved in polynomial time using interior point methods [30]. While our focus in this paper is not on algorithms, we remark that for the structured semidefinite programs discussed in this paper, many different special-purpose methods have been devised.

The main result about semidefinite programming that we use is the following optimality condition (see [30] for example).

THEOREM 2.1. *Suppose* (2.1) *and* (2.2) *are strictly feasible. Then* $X^\star$ *and* $(y^\star, S^\star)$ *are optimal for the primal* (2.1) *and dual* (2.2) *respectively if and only if* $X^\star$ *is primal feasible,* $(y^\star, S^\star)$ *is dual feasible and* $X^\star S^\star = 0$.

2.3. Tractable diagonal and low-rank matrix decompositions. To decompose X into a diagonal part and a positive semidefinite low-rank part, we may try to solve the following rank minimization problem

$$\underset{D,L}{\text{minimize}}\ \mathrm{rank}(L) \quad \text{subject to} \quad \begin{cases} X = D + L \\ L \succeq 0 \\ D \ \text{ diagonal.} \end{cases}$$

Since the rank function is non-convex and non-differentiable, it is not clear how to solve this optimization problem directly. One approach that has been successful for

other rank minimization problems (for example those in [22, 23]), is to replace the rank function with the trace function in the objective. This can be viewed as a convexification of the problem as the trace function is the convex envelope of the rank function when restricted to positive semidefinite matrices with spectral norm at most one. Performing this convexification leads to the semidefinite program we refer to as *minimum trace factor analysis* (MTFA):

$$\underset{D,L}{\text{minimize}}\ \operatorname{tr}(L) \quad \text{subject to} \quad \begin{cases} X = D + L \\ L \succeq 0 \\ D \quad \text{diagonal.} \end{cases} \tag{2.3}$$

It has been shown by Della Riccia and Shapiro [7] that if MTFA is feasible it has a unique optimal solution. One central concern of this paper is to understand when the diagonal and low-rank decomposition of a matrix given by MTFA is 'correct' in the following sense.

Recovery problem I. Suppose X is a matrix of the form $X = D^\star + L^\star$ where $D^\star$ is diagonal and $L^\star$ is positive semidefinite. What conditions on $(D^\star, L^\star)$ ensure that $(D^\star, L^\star)$ is the unique optimum of MTFA with input X?

We establish in Section 3 that whether $(D^\star, L^\star)$ is the unique optimum of MTFA with input $X = D^\star + L^\star$ depends only on the column space of $L^\star$, motivating the following definition.

DEFINITION 2.2. *A subspace $\mathcal{U}$ of $\mathbb{R}^n$ is* recoverable by MTFA *if for every diagonal $D^\star$ and every positive semidefinite $L^\star$ with column space $\mathcal{U}$, $(D^\star, L^\star)$ is the unique optimum of MTFA with input $X = D^\star + L^\star$.*

In these terms, we can restate the recovery problem succinctly as follows.

Recovery problem II. Determine which subspaces of $\mathbb{R}^n$ are recoverable by MTFA.

Much of the basic analysis of MTFA, including optimality conditions and relations between minimum rank and minimum trace factor analysis, was carried out in a sequence of papers by Shapiro [26, 27, 28] and Della Riccia and Shapiro [7]. More recently, Chandrasekaran et al. [6] and Candès et al. [4] considered convex optimization heuristics for decomposing a matrix as a sum of a sparse and low-rank matrix. Since a diagonal matrix is certainly sparse, the analysis in [6] can be specialized to give fairly conservative sufficient conditions for the success of MTFA.

The diagonal and low-rank decomposition problem can also be interpreted as a low-rank matrix completion problem, where we are given all the entries of a low-rank matrix except the diagonal, and aim to correctly reconstruct the diagonal entries. As such, this paper is closely related to the ideas and techniques used in the work of Candès and Recht [5] and a number of subsequent papers on this topic. We would like to emphasize a key point of distinction between that line of work and the present paper. The recent low-rank matrix completion literature largely focuses on determining the proportion of *randomly selected* entries of a low-rank matrix that need to be revealed to be able to reconstruct that low-rank matrix using a tractable algorithm. The results of this paper, on the other hand, can be interpreted as attempting to understand which low-rank matrices can be reconstructed from a *fixed* and quite canonical pattern of revealed entries.

2.4. Faces of the elliptope. The faces of the cone of $n \times n$ positive semidefinite matrices are all of the form

$$\mathcal{F}_{\mathcal{U}} = \{X \succeq 0 : \mathcal{N}(X) \supseteq \mathcal{U}\} \tag{2.4}$$

where $\mathcal{U}$ is a subspace of $\mathbb{R}^n$ [19]. Conversely given any subspace $\mathcal{U}$ of $\mathbb{R}^n$, $\mathcal{F}_\mathcal{U}$ is a face of $\mathcal{S}^n_+$. As a consequence, the faces of $\mathcal{E}_n$ are all of the form

$$\mathcal{E}_n \cap \mathcal{F}_\mathcal{U} = \{X \succeq 0 : \mathcal{N}(X) \supseteq \mathcal{U},\ \mathrm{diag}(X) = \mathbf{1}\} \tag{2.5}$$

where $\mathcal{U}$ is a subspace of $\mathbb{R}^n$ [19]. It is *not* the case, however, that for every subspace $\mathcal{U}$ of $\mathbb{R}^n$ there is a correlation matrix with nullspace containing $\mathcal{U}$, motivating the following definition.

DEFINITION 2.3 ([19]). *A subspace $\mathcal{U}$ of $\mathbb{R}^n$ is* realizable *if there is an $n \times n$ correlation matrix Q such that $\mathcal{N}(Q) \supseteq \mathcal{U}$.*

The problem of understanding the facial structure of the set of correlation matrices can be restated as follows.

Facial structure problem. Determine which subspaces of $\mathbb{R}^n$ are realizable.

Much is already known about the faces of the elliptope. For example, all possible dimensions of faces as well as polyhedral faces, are known [20]. Characterizations of the realizable subspaces of $\mathbb{R}^n$ of dimension 1, $n-2$, and $n-1$ are given in [8] and implicitly in [19] and [20]. Nevertheless, little is known about which k dimensional subspaces of $\mathbb{R}^n$ are realizable for general n and k.

2.5. Ellipsoid fitting.

Ellipsoid fitting problem I. What conditions on a collection of n points in $\mathbb{R}^k$ ensure that there is a centered ellipsoid passing *exactly* through all those points?

Let us consider some basic properties of this problem.

Number of points. If $n \leq k$ we can always fit an ellipsoid to the points. Indeed if V is the matrix with columns $v_1, v_2, \ldots, v_n$ then the image of the unit sphere in $\mathbb{R}^n$ under V is a centered ellipsoid passing through $v_1, v_2, \ldots, v_n$. If $n > \binom{k+1}{2}$ and the points are 'generic' then we cannot fit a centered ellipsoid to them. This is because if we represent the ellipsoid by a symmetric $k \times k$ matrix M, the condition that it passes through the points (ignoring the positivity condition on M) means that M must satisfy n linearly independent equations.

Invariances. If $T \in GL(k)$ is an invertible linear map then there is an ellipsoid passing through $v_1, v_2, \ldots, v_n$ if and only if there is an ellipsoid passing through $Tv_1, Tv_2, \ldots, Tv_n$. This means that whether there is an ellipsoid passing through n points in $\mathbb{R}^k$ does not depend on the actual set of n points, but on a subspace of $\mathbb{R}^n$ related to the points. We summarize this observation in the following lemma.

LEMMA 2.4. *Suppose V is a $k \times n$ matrix with row space $\mathcal{V}$. If there is a centered ellipsoid in $\mathbb{R}^k$ passing through the columns of V then there is a centered ellipsoid passing through the columns of any matrix $\tilde{V}$ with row space $\mathcal{V}$.*

Lemma 2.4 asserts that whether it is possible to fit an ellipsoid to $v_1, v_2, \ldots, v_n$ depends only on the row space of the matrix with columns given by the v_i, motivating the following definition.

DEFINITION 2.5. *A subspace $\mathcal{V}$ of $\mathbb{R}^n$ has the* ellipsoid fitting property *if there is a $k \times n$ matrix V with row space $\mathcal{V}$ and a centered ellipsoid in $\mathbb{R}^k$ that passes through each column of V.*

As such we can restate the ellipsoid fitting problem as follows.

Ellipsoid fitting problem II. Determine which subspaces of $\mathbb{R}^n$ have the ellipsoid fitting property.

3. Relating ellipsoid fitting, diagonal and low-rank decompositions, and correlation matrices. In this section we show that the ellipsoid fitting problem, the recovery problem, and the facial structure problem are equivalent in the following sense.

PROPOSITION 3.1. *Let $\mathcal{U}$ be a subspace of $\mathbb{R}^n$. Then the following are equivalent:*

1. *$\mathcal{U}$ is recoverable by MTFA.*
2. *$\mathcal{U}$ is realizable.*
3. *$\mathcal{U}^\perp$ has the ellipsoid fitting property.*

Proof. To see that 2 implies 3, let V be a $k \times n$ matrix with nullspace $\mathcal{U}$ and let v_i denote the ith column of V. If $\mathcal{U}$ is realizable there is a correlation matrix Y with nullspace containing $\mathcal{U}$. Hence there is some $M \succeq 0$ such that $Y = V^T M V$ and $v_i^T M v_i = 1$ for $i \in [n]$. Since V has nullspace $\mathcal{U}$, it has row space $\mathcal{U}^\perp$. Hence the subspace $\mathcal{U}^\perp$ has the ellipsoid fitting property. By reversing the argument we see that the converse also holds.

The equivalence of 1 and 2 arises from semidefinite programming duality. Following a slight reformulation, MTFA (2.3) can be expressed as

$$\underset{d,L}{\text{maximize}} \ \langle \mathbf{1}, d \rangle \quad \text{subject to} \quad \begin{cases} X = \operatorname{diag}^*(d) + L \\ L \succeq 0 \end{cases} \tag{3.1}$$

and its dual as

$$\underset{Y}{\text{minimize}} \ \langle X, Y \rangle \quad \text{subject to} \quad \begin{cases} \operatorname{diag}(Y) = \mathbf{1} \\ Y \succeq 0 \end{cases} \tag{3.2}$$

which is clearly just the optimization of the linear functional defined by X over the elliptope. We note that (3.1) is exactly in the standard dual form (2.2) for semidefinite programming and correspondingly that (3.2) is in the standard primal form (2.1) for semidefinite programming.

Suppose $\mathcal{U}$ is recoverable by MTFA. Fix a diagonal matrix $D^\star$ and a positive semidefinite matrix $L^\star$ with column space $\mathcal{U}$ and let $X = D^\star + L^\star$. Since (3.1) and (3.2) are strictly feasible, by Theorem 2.1 (optimality conditions for semidefinite programming), the pair $(\operatorname{diag}(D^\star), L^\star)$ is an optimum of (3.1) if and only if there is some correlation matrix $Y^\star$ such that $Y^\star L^\star = 0$. Since $\mathcal{R}(L^\star) = \mathcal{U}$ this implies that $\mathcal{U}$ is realizable. Conversely, if $\mathcal{U}$ is realizable, there is some $Y^\star$ such that $Y^\star L^\star = 0$ for every $L^\star$ with column space $\mathcal{U}$, showing that $\mathcal{U}$ is recoverable by MTFA. □

Remark. We note that in the proof of Proposition 3.1 we established that the two versions of the recovery problem stated in Section 2.3 are actually equivalent. In particular, whether $(D^\star, L^\star)$ is the optimum of MTFA with input $X = D^\star + L^\star$ depends only on the column space of $L^\star$.

3.1. Certificates of failure. We can prove that a subspace $\mathcal{U}$ is realizable by constructing a correlation matrix with nullspace containing $\mathcal{U}$. We can prove that a subspace is *not* realizable by constructing a matrix that *certifies* this fact. Geometrically, a subspace $\mathcal{U}$ is realizable if and only if the subspace $\mathcal{L}_\mathcal{U} = \{X \in \mathcal{S}^n : \mathcal{N}(X) \supseteq \mathcal{U}\}$ of symmetric matrices intersects with the elliptope. So a certificate that $\mathcal{U}$ is not realizable is a hyperplane in the space of symmetric matrices that strictly separates the elliptope from $\mathcal{L}_\mathcal{U}$. The following lemma describes the structure of these separating hyperplanes.

LEMMA 3.2. *A subspace $\mathcal{U}$ of $\mathbb{R}^n$ is* not *realizable if and only if there is a diagonal matrix D such that $tr(D) > 0$ and $v^T D v \leq 0$ for all $v \in \mathcal{U}^\perp$.*

Proof. By Proposition 3.1, $\mathcal{U}$ is not realizable if and only if $\mathcal{U}^\perp$ does not have the ellipsoid fitting property. Let V be a $k \times n$ matrix with row space $\mathcal{U}^\perp$. Then $\mathcal{U}^\perp$ does not have the ellipsoid fitting property if and only if we cannot find an ellipsoid

passing through the columns of V, i.e. the semidefinite program

$$\underset{M}{\text{minimize}}\ \langle 0, M\rangle \quad \text{subject to} \quad \begin{cases} \operatorname{diag}(V^T M V) = \mathbf{1} \\ M \succeq 0 \end{cases} \tag{3.3}$$

is infeasible. The semidefinite programming dual of (3.3) is

$$\underset{d}{\text{maximize}}\ \langle d, \mathbf{1}\rangle \quad \text{subject to} \quad \left\{ \ V \operatorname{diag}^*(d) V^T \preceq 0. \right. \tag{3.4}$$

Since (3.4) is clearly always feasible, by strong duality (which holds because both primal and dual problems are strictly feasible) (3.3) is infeasible if and only if (3.4) is unbounded. This occurs if and only if there is some d with $\sum_{i\in[n]} d_i > 0$ and yet $V \operatorname{diag}^*(d) V^T \preceq 0$. Then $D = \operatorname{diag}^*(d)$ has the properties in the statement of the lemma. □

3.2. Exploiting connections: results for one dimensional subspaces. In 1940, Ledermann [21] characterized the one dimensional subspaces that are recoverable by MTFA. In 1990, Grone et al. [15] gave a necessary condition for a subspace to be realizable. In 1993, independently of Ledermann's work, Delorme and Poljak [8] showed that this condition is also sufficient for one dimensional subspaces. Since we have established that a subspace is recoverable by MTFA if and only if it is realizable, Ledermann's result and Delorme and Poljak's results are equivalent. In this section we translate these equivalent results into the context of the ellipsoid fitting problem, giving a geometric characterization of when it is possible to fit a centered ellipsoid to $k+1$ points in $\mathbb{R}^k$.

Delorme and Poljak state their result in terms of the following definition.

DEFINITION 3.3 ([8]). *A vector $u \in \mathbb{R}^n$ is* balanced *if, for all $i \in [n]$,*

$$|u_i| \leq \sum_{j \neq i} |u_j|. \tag{3.5}$$

If the inequality is strict we say that u is strictly balanced.

In the following, the necessary condition is due to Grone et al. [15] and the sufficient condition is due to Ledermann [21] (in the context of the analysis of MTFA) and Delorme and Poljak [8] (in the context of the facial structure of the elliptope). We state the result only in terms of realizability of a subspace.

THEOREM 3.4. *If a subspace $\mathcal{U}$ of $\mathbb{R}^n$ is realizable then every $u \in \mathcal{U}$ is balanced. If $\mathcal{U} = \operatorname{span}\{u\}$ is one-dimensional then $\mathcal{U}$ is realizable if and only if u is balanced.*

The balance condition has a particularly natural geometric interpretation in the ellipsoid fitting setting (Lemma 3.5, below). The proof is a fairly straightforward application of linear programming duality, which we defer to Appendix A.

LEMMA 3.5. *Suppose V is any $k \times n$ matrix with $\mathcal{N}(V) = \mathcal{U}$. Denote the columns of V by $v_1, v_2, \ldots, v_n \in \mathbb{R}^k$. Then every $u \in \mathcal{U}$ is balanced if and only if for each $i \in [n]$, v_i lies on the boundary of the convex hull of $\pm v_1, \pm v_2, \ldots, \pm v_n$.*

By combining Theorem 3.4 with Lemma 3.5, we are in a position to interpret Theorem 3.4 purely in terms of ellipsoid fitting.

COROLLARY 3.6. *If there is an ellipsoid passing through $\pm v_1, \pm v_2, \ldots, \pm v_n \in \mathbb{R}^k$ then $\pm v_1, \pm v_2, \ldots, \pm v_n$ lie on the boundary of their convex hull. If, in addition, $k = n-1$ the converse also holds.*

We note that $\pm v_1, \pm v_2, \ldots, \pm v_n$ lie on the boundary of their convex hull if and only if there exists *some* convex set with boundary containing $\pm v_1, \pm v_2, \ldots, \pm v_n$. In

this geometric setting, it is clear that this is a necessary condition to be able to find a centered ellipsoid passing through the points, but not so obvious that it is sufficient if $k = n - 1$.

4. A sufficient condition for the three problems. In this section we establish a new sufficient condition for a subspace $\mathcal{U}$ of $\mathbb{R}^n$ to be realizable and consequently a sufficient condition for $\mathcal{U}$ to be recoverable by MTFA and $\mathcal{U}^\perp$ to have the ellipsoid fitting property. Our condition is based on a simple property of a subspace known as coherence.

Given a subspace $\mathcal{U}$ of $\mathbb{R}^n$, the coherence of $\mathcal{U}$ is a measure of how close the subspace is to containing any of the elementary unit vectors. This notion was introduced (with a different scaling) by Candès and Recht in their work on low-rank matrix completion [5], although related quantities have played an important role in the analysis of sparse reconstruction problems since the work of Donoho and Huo [10].

DEFINITION 4.1. *If $\mathcal{U}$ is a subspace of $\mathbb{R}^n$ then the* coherence *of $\mathcal{U}$ is*

$$\mu(\mathcal{U}) = \max_{i \in [n]} \|P_{\mathcal{U}} e_i\|_2^2.$$

A basic property of coherence is that it satisfies the inequality

$$\frac{\dim(\mathcal{U})}{n} \leq \mu(\mathcal{U}) \leq 1 \tag{4.1}$$

for any subspace $\mathcal{U}$ of $\mathbb{R}^n$ [5]. This inequality, together with the definition of coherence, provides useful intuition about the properties of subspaces with low coherence, that is *in*coherence. Any subspace with low coherence is necessarily of low dimension and far from containing any of the elementary unit vectors e_i. As such, any symmetric matrix with incoherent row/column spaces is necessarily of low-rank and quite different from being a diagonal matrix.

4.1. Coherence-threshold-type sufficient conditions. In this section we focus on finding the largest possible α such that

$$\mu(\mathcal{U}) < \alpha \implies \mathcal{U} \text{ is realizable,}$$

that is finding the best possible coherence-threshold-type sufficient condition for a subspace to be realizable. Such conditions are of particular interest because the dependence they have on the ambient dimension and the dimension of the subspace is only the mild dependence implied by (4.1). In contrast, existing results (e.g. [8, 20, 19]) about realizability of subspaces hold only for specific combinations of the ambient dimension and the dimension of the subspace.

The following theorem, our main result, gives a sufficient condition for realizability based on a coherence-threshold condition. Furthermore, it establishes that this is the best possible coherence-threshold-type sufficient condition.

THEOREM 4.2. *If $\mathcal{U}$ is a subspace of $\mathbb{R}^n$ and $\mu(\mathcal{U}) < 1/2$ then $\mathcal{U}$ is realizable. On the other hand, given any $\alpha > 1/2$, there is a subspace $\mathcal{U}$ with $\mu(\mathcal{U}) = \alpha$ that is not realizable.*

Proof. We give the main idea of the proof, deferring some details to Appendix A. Instead of proving that there is some $Y \in \mathcal{F}_{\mathcal{U}} = \{Y \succeq 0 : \mathcal{N}(Y) \supseteq \mathcal{U}\}$ such that $Y_{ii} = 1$ for $i \in [n]$, it suffices to choose a convex cone $\mathcal{K}$ that is an inner approximation to $\mathcal{F}_{\mathcal{U}}$ and establish that there is some $Y \in \mathcal{K}$ such that $Y_{ii} = 1$ for $i \in [n]$. One natural

choice is to take $\mathcal{K} = \{P_{\mathcal{U}^\perp} \operatorname{diag}^*(\lambda) P_{\mathcal{U}^\perp} : \lambda \geq 0\}$, which is clearly contained in $\mathcal{F}_{\mathcal{U}}$. Note that there is some $Y \in \mathcal{K}$ such that $Y_{ii} = 1$ for all $i \in [n]$ if and only if there is $\lambda \geq 0$ such that

$$\operatorname{diag}\left(P_{\mathcal{U}^\perp} \operatorname{diag}^*(\lambda) P_{\mathcal{U}^\perp}\right) = \mathbf{1}. \tag{4.2}$$

The rest of the proof of the sufficient condition involves showing that if $\mu(\mathcal{U}) < 1/2$ then such a non-negative λ exists. We establish this in Lemma A.1.

Now let us construct, for any $\alpha > 1/2$, a subspace with coherence α that is not realizable. Let $\mathcal{U}$ to be the subspace of $\mathbb{R}^2$ spanned by $u = (\sqrt{\alpha}, \sqrt{1-\alpha})$. Then $\mu(\mathcal{U}) = \max\{\alpha, 1-\alpha\} = \alpha$ and yet by Theorem 3.4, $\mathcal{U}$ is not realizable because u is not balanced. □

Remarks. Theorem 4.2 illustrates both the power and limitations of coherence-threshold-type conditions. On the one hand, since coherence is quite a coarse property of a subspace, the result applies to 'many' subspaces (see Proposition 4.6 in Section 4.3). On the other hand, since coherence has very mild dimension dependence, the power of coherence-threshold-type conditions is limited to their specialization to low-dimensional situations, such as one dimensional subspaces of $\mathbb{R}^2$.

4.2. Interpretations of Theorem 4.2. We now establish two corollaries of our coherence-threshold-type sufficient condition for realizability. These corollaries can be thought of as re-interpretations of the coherence inequality $\mu(\mathcal{U}) < 1/2$ in terms of other natural quantities.

An ellipsoid-fitting interpretation. With the aid of Proposition 3.1 we reinterpret our coherence-threshold-type sufficient condition as a sufficient condition on a set of points in $\mathbb{R}^k$ that ensures there is a centered ellipsoid passing through them. The condition involves 'sandwiching' the points between two ellipsoids (that depend on the points). Indeed, given $v_1, v_2, \ldots, v_n \in \mathbb{R}^k$ and $0 < \beta < 1$ we define the ellipsoid

$$\mathcal{E}_\beta(v_1, \ldots, v_n) = \{x \in \mathbb{R}^k : x^T (\textstyle\sum_{j=1}^n v_j v_j^T)^{-1} x \leq \beta\}.$$

DEFINITION 4.3. *Given $0 < \beta < 1$ the points $v_1, v_2, \ldots, v_n$ satisfy the* β-sandwich condition *if*

$$\{v_1, v_2, \ldots, v_n\} \subset \mathcal{E}_1(v_1, \ldots, v_n) \setminus \mathcal{E}_\beta(v_1, \ldots, v_n).$$

The intuition behind this definition (illustrated in Figure 4.1) is that if the points satisfy the β-sandwich condition for β close to one, then they are confined to a thin elliptical shell that is adapted to their position. One might expect that it is 'easier' to fit an ellipsoid to points that are confined in this way. Indeed this is the case.

COROLLARY 4.4. *If $v_1, v_2, \ldots, v_n \in \mathbb{R}^k$ satisfy the $1/2$-sandwich condition then then there is a centered ellipsoid passing through $v_1, v_2, \ldots, v_n$.*

Proof. Let V be the $k \times n$ matrix with columns given by the v_i, and let $\mathcal{U}$ be the nullspace of V. Then the orthogonal projection onto the row space of V is $P_{\mathcal{U}^\perp}$, and can be written as

$$P_{\mathcal{U}^\perp} = V^T (V V^T)^{-1} V.$$

Our assumption that the points satisfy the $1/2$-sandwich condition is equivalent to assuming that $1/2 < [P_{\mathcal{U}^\perp}]_{ii} \leq 1$ for all $i \in [n]$ or alternatively that

$$\mu(\mathcal{U}) = \max_{i \in [n]} [P_{\mathcal{U}}]_{ii} = 1 - \min_{i \in [n]} [P_{\mathcal{U}^\perp}]_{ii} < 1/2.$$

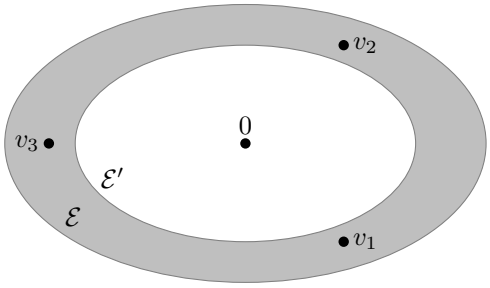


Fig. 4.1: The ellipsoids shown are $\mathcal{E} = \mathcal{E}_1(v_1, v_2, v_3)$ and $\mathcal{E}' = \mathcal{E}_{1/2}(v_1, v_2, v_3)$. There is an ellipsoid passing through v_1, v_2 and v_3 because the points are sandwiched between $\mathcal{E}$ and $\mathcal{E}'$.

From Theorem 4.2 we know that $\mu(\mathcal{U}) < 1/2$ implies that $\mathcal{U}$ is realizable. Invoking Proposition 3.1 we then conclude that there is a centered ellipsoid passing through $v_1, v_2, \ldots, v_n$. □

A balance interpretation. In Section 3.2 we saw that if a subspace $\mathcal{U}$ is realizable, every $u \in \mathcal{U}$ is balanced. The sufficient condition of Theorem 4.2 can be expressed in terms of a balance condition on the element-wise square of the elements of a subspace. (In what follows $u \circ u$ denotes the element-wise square of a vector in $\mathbb{R}^n$.)

COROLLARY 4.5. *Suppose $\mathcal{U}$ is a subspace of $\mathbb{R}^n$. If $u \circ u$ is strictly balanced for every $u \in \mathcal{U}$ then $\mathcal{U}$ is realizable.*

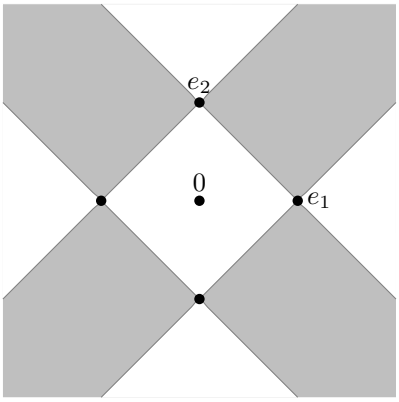
Proof. It suffices to show that if for every $u \in \mathcal{U}$, $u \circ u$ is strictly balanced, then $\mu(\mathcal{U}) < 1/2$ (although we could reverse the argument to establish the equivalence of these conditions). If $u \circ u$ is strictly balanced for all $u \in \mathcal{U}$ then for all $i \in [n]$ and all $u \in \mathcal{U}$

$$2\langle e_i, u\rangle^2 < \sum_{j=1}^{n} \langle e_j, u\rangle^2 = \|u\|_2^2. \tag{4.3}$$

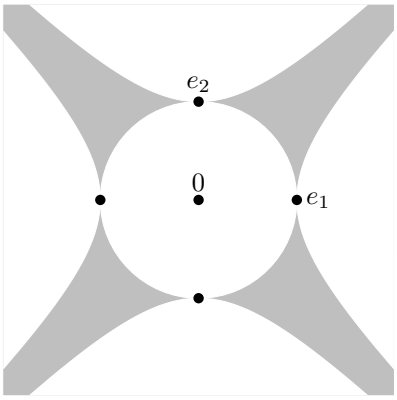
Since $\|P_{\mathcal{U}} e_i\|_2 = \max_{u \in \mathcal{U}\setminus\{0\}} \langle e_i, u\rangle / \|u\|_2$, it follows from (4.3) that $2\|P_{\mathcal{U}} e_i\|_2^2 < 1$. Since this holds for all $i \in [n]$ it follows that $\mu(\mathcal{U}) < 1/2$. □

Remark. Suppose $\mathcal{U} = \text{span}\{u\}$ is a one-dimensional subspace of $\mathbb{R}^n$. We have just established that if $u \circ u$ is strictly balanced then $\mathcal{U}$ is realizable and so (by Theorem 3.4) u must be balanced. We note that it is straightforward to establish directly that if $u \circ u$ is balanced then u is balanced by using the definition of balance and the fact that $\|x\|_1 \geq \|x\|_2$ for any $x \in \mathbb{R}^n$.

4.3. Examples. To gain more intuition for what Theorem 4.2 means, we consider its implications in two particular cases. First, we compare the characterization of when it is possible to fit an ellipsoid to $k+1$ points in $\mathbb{R}^k$ (Corollary 3.6) with the specialization of our sufficient condition to this case (Corollary 4.4). This comparison provides some insight into how conservative our sufficient condition is. Second, we investigate the coherence properties of suitably 'random' subspaces. This provides intuition about whether or not $\mu(\mathcal{U}) < 1/2$ is a very restrictive condition. In particular, we establish that 'most' subspaces of $\mathbb{R}^n$ with dimension bounded above by $(1/2 - \epsilon)n$ are realizable.



(a) The shaded set is R, those points v for which we can fit an ellipsoid through v and the standard basis vectors.



(b) The shaded set is R', those points v such that v, e_1 and e_2 satisify the condition of Corollary 4.4.

Fig. 4.2: Comparing our sufficient condition for ellipsoid fitting (Corollary 4.4) with the characterization (Corollary 3.6) in the case of fitting an ellipsoid to $k+1$ points in $\mathbb{R}^k$.

Fitting an ellipsoid to $k+1$ points in $\mathbb{R}^k$. Recall that Ledermann and Delorme and Poljak's result, interpreted in terms of ellipsoid fitting, tells us that we can fit an ellipsoid to $k+1$ points $v_1, \ldots, v_{k+1} \in \mathbb{R}^k$ if and only if those points are on the boundary of the convex hull of $\{\pm v_1, \ldots, \pm v_{k+1}\}$ (see Corollary 3.6). We now compare this characterization with the 1/2-sandwich condition, which is sufficient by Corollary 4.4.

Without loss of generality we assume that k of the points are $e_1, \ldots, e_k$, the standard basis vectors, and compare the conditions by considering the set of locations of the $k+1$st point $v \in \mathbb{R}^k$ for which we can fit an ellipsoid through all $k+1$ points. Corollary 3.6 gives a characterization of this region as

$$R = \{v \in \mathbb{R}^k : \sum_{j=1}^{k} |v_j| \geq 1,\ |v_i| - \sum_{j \neq i} |v_j| \leq 1 \quad \text{for } i \in [k]\}$$

which is shown in Figure 4.2a in the case $k = 2$. The set of v such that $v, e_1, \ldots, e_n$ satisfy the 1/2-sandwich condition can be written as

$$\begin{aligned} R' &= \{v \in \mathbb{R}^k : v^T (I + vv^T)^{-1} v > 1/2,\ e_i^T (I + vv^T)^{-1} e_i > 1/2 \quad \text{for } i \in [k]\} \\ &= \{v \in \mathbb{R}^k : \sum_{j=1}^{k} v_j^2 > 1,\ v_i^2 - \sum_{j \neq i} v_j^2 < 1 \quad \text{for } i \in [k]\} \end{aligned}$$

which is shown in Figure 4.2b. It is clear that $R' \subseteq R$.

Realizability of random subspaces. Suppose $\mathcal{U}$ is a subspace generated by taking the column space of an $n \times r$ matrix with i.i.d. standard Gaussian entries. For what values of r and n does such a subspace have $\mu(\mathcal{U}) < 1/2$ with high probability, i.e. satisfy our sufficient condition for being realizable?

The following result essentially shows that for large n, 'most' subspaces of dimension at most $(1/2-\epsilon)n$ are realizable. This suggests that MTFA is a very good heuristic for diagonal and low-rank decomposition problems in the high-dimensional setting. Indeed 'most' subspaces of dimension up to one half the ambient dimension—hardly just low-dimensional subspaces—are recoverable by MTFA.

PROPOSITION 4.6. *Let $0 < \epsilon < 1/2$ be a constant and suppose $n > 6/(\epsilon^2 - 2\epsilon^3)$. There are positive constants $\bar{c}$, $\tilde{c}$, (depending only on ϵ) such that if $\mathcal{U}$ is a random $(1/2-\epsilon)n$ dimensional subspace of $\mathbb{R}^n$ then*

$$\Pr[\mathcal{U} \textit{ is realizable}] \geq 1 - \bar{c}\sqrt{n}e^{-\tilde{c}n}.$$

We provide a proof of this result in Appendix A. The main idea is that the coherence of a random r dimensional subspace of $\mathbb{R}^n$ is the maximum of n random variables that concentrate around their mean of r/n for large n.

To illustrate the result, we consider the case where $\epsilon = 1/4$ and $n > 192$. Then (by examining the proof in Appendix A) we see that we can take $\tilde{c} = 1/24$ and $\bar{c} = 24/\sqrt{3\pi} \approx 7.8$. Hence if $n > 192$ and $\mathcal{U}$ is a random $n/4$ dimensional subspace of $\mathbb{R}^n$ we have that

$$\Pr[\mathcal{U} \text{ is realizable}] \geq 1 - 7.8\sqrt{n}e^{-n/24}.$$

5. Tractable block diagonal and low-rank decompositions and related problems. In this section we generalize our results to the analogue of MTFA for *block*-diagonal and low-rank decompositions. Mimicking our earlier development, we relate the analysis of this variant of MTFA to the facial structure of a variant of the elliptope and a generalization of the ellipsoid fitting problem. The key point is that these problems all possess additional symmetries that, once taken into account, essentially allow us to reduce our analysis to cases already considered in Sections 3 and 4.

Throughout this section, let $\mathcal{P}$ be a fixed partition of $\{1, 2, \ldots, n\}$. We say a matrix is $\mathcal{P}$-block-diagonal if it is zero except for the principal submatrices indexed by the elements of $\mathcal{P}$. We denote by $\text{blkdiag}_{\mathcal{P}}$ the map that takes an $n \times n$ matrix and maps it to the principal submatrices indexed by $\mathcal{P}$. Its adjoint, denoted $\text{blkdiag}^*_{\mathcal{P}}$, takes a tuple of symmetric matrices $(X_{\mathcal{I}})_{\mathcal{I}\in\mathcal{P}}$ and produces an $n \times n$ matrix that is $\mathcal{P}$-block diagonal with blocks given by the $X_{\mathcal{I}}$.

We now describe the analogues of MTFA, ellipsoid fitting, and the problem of determining the facial structure of the elliptope.

Block minimum trace factor analysis. If $X = B^\star + L^\star$ where $B^\star$ is $\mathcal{P}$-block-diagonal and $L^\star \succeq 0$ is low rank, the obvious analogue of MTFA is the semidefinite program

$$\underset{B,L}{\text{minimize}}\ \text{tr}(L) \quad \text{subject to} \quad \begin{cases} X = B + L \\ L \ \succeq \ 0 \\ B \text{ is } \mathcal{P}\text{-block-diagonal} \end{cases} \tag{5.1}$$

which we call *block minimum trace factor analysis* (BMTFA).

DEFINITION 5.1. *A subspace $\mathcal{U}$ of $\mathbb{R}^n$ is* recoverable by BMTFA *if for every $B^\star$ that is $\mathcal{P}$-block-diagonal and every positive semidefinite $L^\star$ with column space $\mathcal{U}$, $(B^\star, L^\star)$ is the unique optimum of BMTFA with input $X = B^\star + L^\star$.*

Faces of the $\mathcal{P}$-elliptope. Just as MTFA is related to the facial structure of the elliptope, BMTFA is related to the facial structure of the spectrahedron

$$\mathcal{E}_{\mathcal{P}} = \{Y \succeq 0 : \text{blkdiag}_{\mathcal{P}}(Y) = (I, I, \ldots, I)\}.$$

We refer to $\mathcal{E}_{\mathcal{P}}$ as the $\mathcal{P}$-elliptope. We extend the definition of a realizable subspace to this context.

DEFINITION 5.2. *A subspace $\mathcal{U}$ of $\mathbb{R}^n$ is* $\mathcal{P}$-realizable *if there is some $Y \in \mathcal{E}_{\mathcal{P}}$ such that $\mathcal{N}(Y) \supseteq \mathcal{U}$.*

Generalized ellipsoid fitting. To describe the $\mathcal{P}$-ellipsoid fitting problem we first introduce some convenient notation. If $\mathcal{I} \subset [n]$ we write

$$S^{\mathcal{I}} = \{x \in \mathbb{R}^n : \|x\|_2 = 1,\ x_j = 0 \text{ if } j \notin \mathcal{I}\} \tag{5.2}$$

for the intersection of the unit sphere with the coordinate subspace indexed by $\mathcal{I}$.

Suppose $v_1, v_2, \ldots, v_n \in \mathbb{R}^k$ is a collection of points and V is the $k \times n$ matrix with columns given by the v_i. Noting that $S^{\{i\}} = \{-e_i, e_i\}$, and thinking of V as a linear map from $\mathbb{R}^n$ to $\mathbb{R}^k$, we see that the ellipsoid fitting problem is to find an ellipsoid in $\mathbb{R}^k$ with boundary containing $\cup_{i\in[n]} V(S^{\{i\}})$, i.e. the collection of points $\pm v_1, \ldots, \pm v_n$. The $\mathcal{P}$-ellipsoid fitting problem is then to find an ellipsoid in $\mathbb{R}^k$ with boundary containing $\cup_{\mathcal{I}\in\mathcal{P}} V(S^{\mathcal{I}})$, i.e. the collection of ellipsoids $V(S^{\mathcal{I}})$.

The generalization of the ellipsoid fitting property of a subspace is as follows.

DEFINITION 5.3. *A subspace $\mathcal{V}$ of $\mathbb{R}^n$ has the* $\mathcal{P}$-ellipsoid fitting property *if there is a $k \times n$ matrix V with row space $\mathcal{V}$ such that there is a centered ellipsoid in $\mathbb{R}^k$ with boundary containing $\cup_{\mathcal{I}\in\mathcal{P}} V(S^{\mathcal{I}})$.*

5.1. Relating the generalized problems. The facial structure of the $\mathcal{P}$-elliptope, BMTFA, and the $\mathcal{P}$-ellipsoid fitting problem are related by the following result, the proof of which is omitted as it is almost identical to that of Proposition 3.1.

PROPOSITION 5.4. *Let $\mathcal{U}$ be a subspace of $\mathbb{R}^n$. Then the following are equivalent:*

1. *$\mathcal{U}$ is recoverable by BMTFA.*
2. *$\mathcal{U}$ is $\mathcal{P}$-realizable.*
3. *$\mathcal{U}^\perp$ has the $\mathcal{P}$-ellipsoid fitting property.*

The following lemma is the analogue of Lemma 3.2. It describes certificates that a subspace $\mathcal{U}$ is not $\mathcal{P}$-realizable. Again the proof is almost identical to that of Lemma 3.2 so we omit it.

LEMMA 5.5. *A subspace $\mathcal{U}$ of $\mathbb{R}^n$ is* not *$\mathcal{P}$-realizable if and only if there is a $\mathcal{P}$-block-diagonal matrix B such that $tr(B) > 0$ and $v^T B v \leq 0$ for all $v \in \mathcal{U}^\perp$.*

For the sake of brevity, in what follows we only discuss the problem of whether $\mathcal{U}$ is $\mathcal{P}$-realizable without explicitly translating the results into the context of the other two problems.

5.2. Symmetries of the $\mathcal{P}$-elliptope. We now consider the symmetries of the $\mathcal{P}$-elliptope. Our motivation for doing so is that it allows us to partition subspaces into classes for which either all elements are $\mathcal{P}$-realizable or none of the elements are $\mathcal{P}$-realizable.

It is clear that the $\mathcal{P}$-elliptope is invariant under conjugation by $\mathcal{P}$-block-diagonal orthogonal matrices. Let $G_{\mathcal{P}}$ denote this subgroup of the group of $n \times n$ orthogonal matrices. There is a natural action of $G_{\mathcal{P}}$ on subspaces of $\mathbb{R}^n$ defined as follows. If $P \in G_{\mathcal{P}}$ and $\mathcal{U}$ is a subspace of $\mathbb{R}^n$ then $P \cdot \mathcal{U}$ is the image of the subspace $\mathcal{U}$ under

the map P. (It is straightforward to check that this is a well defined group action.) If there exists some $P \in G_{\mathcal{P}}$ such that $P \cdot \mathcal{U} = \mathcal{U}'$ then we write $\mathcal{U} \sim \mathcal{U}'$ and say that $\mathcal{U}$ and $\mathcal{U}'$ are *equivalent*. We care about this equivalence relation on subspaces because the property of being $\mathcal{P}$-realizable is really a property of the corresponding equivalence classes.

PROPOSITION 5.6. *Suppose $\mathcal{U}$ and $\mathcal{U}'$ are subspaces of $\mathbb{R}^n$. If $\mathcal{U} \sim \mathcal{U}'$ then $\mathcal{U}$ is $\mathcal{P}$-realizable if and only if $\mathcal{U}'$ is $\mathcal{P}$-realizable.*

Proof. If $\mathcal{U}$ is $\mathcal{P}$-realizable there is $Y \in \mathcal{E}_{\mathcal{P}}$ such that $Yu = 0$ for all $u \in \mathcal{U}$. Suppose $\mathcal{U}' = P \cdot \mathcal{U}$ for some $P \in G_{\mathcal{P}}$ and let $Y' = PYP^T$. Then $Y' \in \mathcal{E}_{\mathcal{P}}$ and $Y'(Pu) = (PYP^T)(Pu) = 0$ for all $u \in \mathcal{U}$. By the definition of $\mathcal{U}'$ it is then the case that $Y'u' = 0$ for all $u' \in \mathcal{U}'$. Hence $\mathcal{U}'$ is $\mathcal{P}$-realizable. The converse clearly also holds. ∎

5.3. Exploiting symmetries: relating realizability and $\mathcal{P}$-realizability. For a subspace of $\mathbb{R}^n$, we now consider how the notions of $\mathcal{P}$-realizability and realizability (i.e. $[n]$-realizability) relate to each other. Since $\mathcal{E}_{\mathcal{P}} \subset \mathcal{E}_n$, if $\mathcal{U}$ is $\mathcal{P}$-realizable, it is certainly also realizable. While the converse does not hold, we can establish the following partial converse, which we subsequently use to extend our analysis from Sections 3 and 4 to the present setting.

THEOREM 5.7. *A subspace $\mathcal{U}$ of $\mathbb{R}^n$ is $\mathcal{P}$-realizable if and only if $\mathcal{U}'$ is realizable for every $\mathcal{U}'$ such that $\mathcal{U}' \sim \mathcal{U}$.*

Proof. We note that one direction of the proof is obvious since $\mathcal{P}$-realizability implies realizability. It remains to show that if $\mathcal{U}$ is not $\mathcal{P}$-realizable then there is some $\mathcal{U}'$ equivalent to $\mathcal{U}$ that is not realizable.

Recall from Lemma 5.5 that if $\mathcal{U}$ is not $\mathcal{P}$-realizable there is some $\mathcal{P}$-block-diagonal X with positive trace such that $v^T X v \leq 0$ for all $v \in \mathcal{U}^\perp$. Since X is $\mathcal{P}$-block-diagonal there is some $P \in G_{\mathcal{P}}$ such that PXP^T is diagonal. Since conjugation by orthogonal matrices preserves eigenvalues, $\operatorname{tr}(PXP^T) = \operatorname{tr}(X) > 0$. Furthermore $v^T(PXP^T)v = (P^T v)^T X (P^T v) \leq 0$ for all $P^T v \in \mathcal{U}^\perp$. Hence $w^T(PXP^T)w \geq 0$ for all $w \in P \cdot \mathcal{U}^\perp = (P \cdot \mathcal{U})^\perp$. By Lemma 3.2, PXP^T is a certificate that $P \cdot \mathcal{U}$ is not realizable, completing the proof. ∎

The power of Theorem 5.7 lies in its ability to turn any condition for a subspace to be realizable into a condition for the subspace to be $\mathcal{P}$-realizable by appropriately symmetrizing the condition with respect to the action of $G_{\mathcal{P}}$. We now illustrate this approach by generalizing Theorem 3.4 and our coherence based condition (Theorem 4.2) for a subspace to be $\mathcal{P}$-realizable. In each case we first define an appropriately symmetrized version of the original condition. The natural symmetrized version of the notion of balance is as follows.

DEFINITION 5.8. *A vector $u \in \mathbb{R}^n$ is* $\mathcal{P}$-balanced *if for all $\mathcal{I} \in \mathcal{P}$*

$$\|u_{\mathcal{I}}\|_2 \leq \sum_{\mathcal{J} \in \mathcal{P} \setminus \{\mathcal{I}\}} \|u_{\mathcal{J}}\|_2.$$

We next define the appropriately symmetrized analogue of coherence. Just as coherence measures how far a subspace is from any one-dimensional coordinate subspace, $\mathcal{P}$-coherence measures how far a subspace is from any of the coordinate subspaces indexed by elements of $\mathcal{P}$.

DEFINITION 5.9. *The $\mathcal{P}$-coherence of a subspace $\mathcal{U}$ of $\mathbb{R}^n$ is*

$$\mu_{\mathcal{P}}(\mathcal{U}) = \max_{\mathcal{I} \in \mathcal{P}} \max_{x \in S^{\mathcal{I}}} \|P_{\mathcal{U}} x\|_2^2.$$

Just as the coherence of $\mathcal{U}$ can be computed by taking the maximum diagonal element of $P_{\mathcal{U}}$, it is straightforward to veify that the $\mathcal{P}$-coherence of $\mathcal{U}$ can be computed by taking the maximum of the spectral norms of the principal submatrices $[P_{\mathcal{U}}]_{\mathcal{I}}$ indexed by $\mathcal{I} \in \mathcal{P}$.

We now use Theorem 5.7 to establish the natural generalization of Theorem 3.4.

COROLLARY 5.10. *If a subspace $\mathcal{U}$ of $\mathbb{R}^n$ is $\mathcal{P}$-realizable then every element of $\mathcal{U}$ is $\mathcal{P}$-balanced. If $\mathcal{U} = \text{span}\{u\}$ is one dimensional then $\mathcal{U}$ is $\mathcal{P}$-realizable if and only if u is $\mathcal{P}$-balanced.*

Proof. If there is $u \in \mathcal{U}$ that is not $\mathcal{P}$-balanced then there is $P \in G_{\mathcal{P}}$ such that Pu is not balanced (choose P so that it rotates each $u_{\mathcal{I}}$ until it has only one non-zero entry). But then $P \cdot \mathcal{U}$ is not realizable and so $\mathcal{U}$ is not $\mathcal{P}$-realizable.

For the converse, we first show that if a vector is $\mathcal{P}$-balanced then it is balanced. Let $\mathcal{I} \in \mathcal{P}$, and consider $i \in \mathcal{I}$. Then since u is $\mathcal{P}$-balanced,

$$2|u_i| \leq 2\|u_{\mathcal{I}}\|_2 \leq \sum_{\mathcal{J} \in \mathcal{P}} \|u_{\mathcal{J}}\|_2 \leq \sum_{i=1}^{n} |u_i|$$

and so u is balanced.

Now suppose $\mathcal{U} = \text{span}\{u\}$ is one dimensional and u is $\mathcal{P}$-balanced. Since u is $\mathcal{P}$-balanced it follows that Pu is $\mathcal{P}$-balanced (and hence balanced) every $P \in G_{\mathcal{P}}$. Then by Theorem 3.4 $\text{span}\{Pu\}$ is realizable for every $P \in G_{\mathcal{P}}$. Hence by Theorem 5.7, $\mathcal{U}$ is $\mathcal{P}$-realizable. □

Similarly, with the aid of Theorem 5.7 we can write down a $\mathcal{P}$-coherence-threshold condition that is a sufficient condition for a subspace to be $\mathcal{P}$-realizable. The following is a natural generalization of Theorem 4.2.

COROLLARY 5.11. *If $\mu_{\mathcal{P}}(\mathcal{U}) < 1/2$ then $\mathcal{U}$ is $\mathcal{P}$-realizable.*

Proof. By examining the constraints in the variational definitions of $\mu(\mathcal{U})$ and $\mu_{\mathcal{P}}(\mathcal{U})$ we see that $\mu(\mathcal{U}) \leq \mu_{\mathcal{P}}(\mathcal{U})$. Consequently if $\mu_{\mathcal{P}}(\mathcal{U}) < 1/2$ it follows from Theorem 4.2 that $\mathcal{U}$ is realizable. Since $\mu_{\mathcal{P}}$ is invariant under the action of $G_{\mathcal{P}}$ on subspaces we can apply Theorem 5.7 to complete the proof. □

6. Conclusions. We established a link between three problems of independent interest: deciding whether there is a centered ellipsoid passing through a collection of points, understanding the structure of the faces of the elliptope, and deciding which pairs of diagonal and low rank-matrices can be recovered from their sum using a tractable semidefinite-programming-based heuristic, namely minimum trace factor analysis. We provided a simple sufficient condition, based on the notion of the coherence of a subspace, which ensures the success of minimum trace factor analysis, and showed that this is the best possible coherence-threshold-type sufficient condition for this problem. We provided natural generalizations of our results to the problem of analyzing tractable block-diagonal and low-rank decompositions, showing how the symmetries of this problem allow us to reduce much of the analysis to the original diagonal and low-rank case.

Our results suggest both the power and the limitations of using 'coarse' properties of a subspace such as coherence to gain understanding of the faces of the elliptope (and related problems). The power of results based on such properties is that they do not have explicit dimension-dependence, unlike previous results on the faces of the elliptope. At the same time, the lack of explicit dimension dependence typically yields

conservative sufficient conditions for high-dimensional problems. It would be interesting to find a hierarchy of coherence-like conditions that provide less conservative sufficient conditions for higher dimensional problem instances.

Appendix A. Additional proofs.

A.1. Proof of Lemma 3.5. We first establish Lemma 3.5 which gives an interpretation of the balance condition in terms of ellipsoid fitting.

Proof. The proof is a fairly straightforward application of linear programming duality. Throughout let V be the $k \times n$ matrix with columns given by the v_i. The point $v_i \in \mathbb{R}^k$ is on the boundary of the convex hull of $\pm v_1, \ldots, \pm v_n$ if and only if there exists $x \in \mathbb{R}^k$ such that $\langle x, v_i \rangle = 1$ and $|\langle x, v_j \rangle| \leq 1$ for all $j \neq i$. Equivalently, the following linear program (which depends on i) is feasible

$$\underset{x}{\text{minimize}} \ \langle 0, x \rangle \quad \text{subject to} \quad \begin{cases} v_i^T x = 1 \\ |v_j^T x| \leq 1 \ \text{ for all } j \neq i. \end{cases} \tag{A.1}$$

Suppose there is some i such that v_i is in the interior of $\text{conv}\{\pm v_1, \ldots, \pm v_n\}$. Then (A.1) is not feasible so the dual linear program (which depends on i)

$$\underset{u}{\text{maximize}} \ u_i - \sum_{j \neq i} |u_j| \quad \text{subject to} \quad Vu = 0 \tag{A.2}$$

is unbounded. This is the case if and only if there is some u in the nullspace of V such that $u_i > \sum_{j \neq i} |u_j|$. If such a u exists, then it is certainly the case that $|u_i| \geq u_i > \sum_{j \neq i} |u_j|$ and so u is not balanced.

Conversely if u is in the nullspace of V and u is not balanced then either u or $-u$ satisfies $u_i > \sum_{j \neq i} |u_j|$ for some i. Hence the linear program (A.2) associated with the index i is unbounded and so the corresponding linear program (A.1) is infeasible. It follows that v_i is in the interior of the convex hull of $\pm v_1, \ldots, \pm v_n$. ∎

A.2. Completing the proof of Theorem 4.2. We now complete the proof of Theorem 4.2 by establishing the following result about the existence of a non-negative solution to the linear system (4.2).

LEMMA A.1. *If $\mu(\mathcal{U}) < 1/2$ then there is $\lambda \geq 0$ such that*

$$\text{diag}\left(P_{\mathcal{U}^\perp} \, \text{diag}^*(\lambda) P_{\mathcal{U}^\perp}\right) = \mathbf{1}. \tag{A.3}$$

Proof. We note that the linear system (A.3) can be written as $P_{\mathcal{U}^\perp} \circ P_{\mathcal{U}^\perp} \lambda = \mathbf{1}$ where $\circ$ denotes the entry-wise product of matrices. As such, we need to show that $P_{\mathcal{U}^\perp} \circ P_{\mathcal{U}^\perp}$ is invertible and $(P_{\mathcal{U}^\perp} \circ P_{\mathcal{U}^\perp})^{-1} \mathbf{1} \geq 0$. To do so, we appeal to the following (slight restatement) of a theorem of Walters [31] regarding positive solutions to certain linear systems.

THEOREM A.2 (Walters [31]). *Suppose A is a square matrix with non-negative entries and positive diagonal entries. Let D be a diagonal matrix with $D_{ii} = A_{ii}$ for all i. If $y > 0$ and $2y - AD^{-1}y > 0$ then A is invertible and $A^{-1}y > 0$.*

In our case we take $A = P_{\mathcal{U}^\perp} \circ P_{\mathcal{U}^\perp}$ and $y = \mathbf{1}$ in Theorem A.2. It is clear that $P_{\mathcal{U}^\perp} \circ P_{\mathcal{U}^\perp}$ is entry-wise non-negative. Furthermore $[P_{\mathcal{U}^\perp}]_{ii} = 1 - [P_{\mathcal{U}}]_{ii} > 1 - \mu(\mathcal{U}) > 1/2$ and so $D_{ii} = [P_{\mathcal{U}^\perp} \circ P_{\mathcal{U}^\perp}]_{ii} > 1/4$. It then remains to show that

$P_{\mathcal{U}^\perp} \circ P_{\mathcal{U}^\perp} D^{-1}\mathbf{1} < 2\mathbf{1}$. Consider the ith such inequality, and observe that

$$\begin{aligned}
[P_{\mathcal{U}^\perp} \circ P_{\mathcal{U}^\perp} D^{-1}\mathbf{1}]_i &= \left(P_{\mathcal{U}^\perp} D^{-1} P_{\mathcal{U}^\perp}\right)_{ii} \\
&= \left(P_{\mathcal{U}^\perp} D_{ii}^{-1} e_i e_i^T P_{\mathcal{U}^\perp}\right)_{ii} + \left(P_{\mathcal{U}^\perp}(D^{-1} - D_{ii}^{-1} e_i e_i^T) P_{\mathcal{U}^\perp}\right)_{ii} \\
&\leq 1 + \max_{j\in[n]} D_{jj}^{-1} \left(P_{\mathcal{U}^\perp}(I - e_i e_i^T) P_{\mathcal{U}^\perp}\right)_{ii} \\
&< 1 + 4[P_{\mathcal{U}^\perp}]_{ii} - 4[P_{\mathcal{U}^\perp}]_{ii}^2 \\
&= 2 - 4([P_{\mathcal{U}^\perp}]_{ii} - 1/2)^2 \\
&\leq 2
\end{aligned}$$

where we have used the assumption that $[P_{\mathcal{U}^\perp}]_{ii} > 1/2$ for all i and the fact that $P_{\mathcal{U}^\perp}^2 = P_{\mathcal{U}^\perp}$. Applying Walters's theorem completes the proof. □

A.3. Proof of Proposition 4.6. We now establish Proposition 4.6, giving a bound on the probability that a suitably random subspace is realizable by bounding the probability that it has coherence strictly bounded above by 1/2.

Proof. It suffices to show that $\|P_{\mathcal{U}} e_i\|^2 \leq (1-2\epsilon)(1/2-\epsilon) = 1/2 - 2\epsilon^2 < 1/2$ for all i with high probability. The main observation we use is that if $\mathcal{U}$ is a random r dimensional subspace of $\mathbb{R}^n$ and x is any fixed vector with $\|x\| = 1$ then $\|P_{\mathcal{U}} x\|^2 \sim \beta(r/2, (n-r)/2)$ where $\beta(p,q)$ denotes the beta distribution [13]. In the case where $r = (1/2-\epsilon)n$, using a tail bound for β random variables [13] we see that if $x \in \mathbb{R}^n$ is fixed and $r > 3/\epsilon^2$ then

$$\Pr[\|P_{\mathcal{U}} x\|^2 \geq (1+2\epsilon)(1/2-\epsilon)] < \frac{1}{a_\epsilon} \frac{1}{(\pi(1/4-\epsilon^2))^{1/2}} n^{-1/2} e^{-a_\epsilon k}$$

where $a_\epsilon = \epsilon - 4\epsilon^2/3$. Taking a union bound over n events, as long as $r > 3/\epsilon^2$

$$\begin{aligned}
\Pr\left[\mu(\mathcal{U}) \geq 1/2\right] &\leq \Pr\left[\|P_{\mathcal{U}} e_i\|^2 \geq (1-2\epsilon)(1/2-\epsilon) \text{ for some } i \in [n]\right] \\
&\leq n \cdot \frac{1}{a_\epsilon(\pi(1/4-\epsilon^2))^{1/2}} n^{-1/2} e^{-a_\epsilon k} = \bar{c} n^{1/2} e^{-\tilde{c} n}
\end{aligned}$$

for appropriate positive constants $\bar{c}$ and $\tilde{c}$. □

Acknowledgements. The authors would like to thank Prof. Sanjoy Mitter for helpful discussions.

REFERENCES

[1] A.A. Albert, *The matrices of factor analysis*, Proc. Natl. Acad. Sci. USA, 30 (1944), p. 90.
[2] J. Briët, F. de Oliveira Filho, and F. Vallentin, *Grothendieck inequalities for semidefinite programs with rank constraint*, Arxiv preprint arXiv:1011.1754, (2010).
[3] A.E. Brouwer and J. Draisma, *Equivariant Gröbner bases and the Gaussian two-factor model*, Math. Comp., 80 (2011), pp. 1123–1133.
[4] E.J. Candès, X. Li, Y. Ma, and J. Wright, *Robust principal component analysis?*, Journal of the ACM, 58 (2011), pp. 11:1–11:37.
[5] E.J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Found. Comput. Math., 9 (2009), pp. 717–772.
[6] V. Chandrasekaran, S. Sanghavi, P.A. Parrilo, and A.S. Willsky, *Rank-sparsity incoherence for matrix decomposition*, SIAM J. Optim., 21 (2011), pp. 572–596.
[7] G. Della Riccia and A. Shapiro, *Minimum rank and minimum trace of covariance matrices*, Psychometrika, 47 (1982), pp. 443–448.
[8] C. Delorme and S. Poljak, *Combinatorial properties and the complexity of a max-cut approximation*, European J. Combin., 14 (1993), pp. 313–333.

[9] A.P. Dempster, N.M. Laird, and D.B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, J. R. Stat. Soc. Ser. B Stat. Methodol., 39 (1977), pp. 1–38.
[10] D.L. Donoho and X. Huo, *Uncertainty principles and ideal atomic decomposition*, IEEE Trans. Inform Theory, 47 (2001), pp. 2845–2862.
[11] M. Drton, B. Sturmfels, and S. Sullivant, *Algebraic factor analysis: tetrads, pentads and beyond*, Probab. Theory Related Fields, 138 (2007), pp. 463–493.
[12] M. Fazel and J. Goodman, *Approximations for partially coherent optical imaging systems*, tech. report, Stanford University, 1998.
[13] P. Frankl and H. Maehara, *Some geometric applications of the beta distribution*, Ann. Inst. Statist. Math., 42 (1990), pp. 463–474.
[14] M.X. Goemans and D.P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. ACM, 42 (1995), pp. 1115–1145.
[15] R. Grone, S. Pierce, and W. Watkins, *Extremal correlation matrices*, Linear Algebra Appl., 134 (1990), pp. 63–70.
[16] N.J. Higham, *Computing the nearest correlation matrix—a problem from finance*, IMA J. Numer. Anal., 22 (2002), pp. 329–343.
[17] R.E. Kalman, *Identification of noisy systems*, Russian Math. Surveys, 40 (1985), pp. 25–42.
[18] H. Krim and M. Viberg, *Two decades of array signal processing research*, IEEE Signal Process. mag., 13 (1996), pp. 67–94.
[19] M. Laurent and S. Poljak, *On a positive semidefinite relaxation of the cut polytope*, Linear Algebra Appl., 223 (1995), pp. 439–461.
[20] ———, *On the facial structure of the set of correlation matrices*, SIAM J. Matrix Anal. Appl., 17 (1995), pp. 530–547.
[21] W. Ledermann, *On a problem concerning matrices with variable diagonal elements*, Proc. Roy. Soc. Edinburgh, 60 (1940), pp. 1–17.
[22] M. Mesbahi and G.P. Papavassilopoulos, *On the rank minimization problem over a positive semidefinite linear matrix inequality*, IEEE Trans. Automat. Control, 42 (1997), pp. 239–243.
[23] B. Recht, M. Fazel, and P.A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501.
[24] J. Saunderson, *Subspace identification via convex optimization*, master's thesis, Massachusetts Institute of Technology, 2011.
[25] P. Shah and V. Chandrasekaran, *Group symmetry and covariance regularization*, Arxiv preprint arXiv:1111.7061, (2011).
[26] A. Shapiro, *Rank-reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis*, Psychometrika, 47 (1982), pp. 187–199.
[27] ———, *Weighted minimum trace factor analysis*, Psychometrika, 47 (1982), pp. 243–264.
[28] ———, *Identifiability of factor analysis: some results and open problems*, Linear Algebra Appl., 70 (1985), pp. 1–7.
[29] C. Spearman, *'General intelligence,' objectively determined and measured*, American J. Psychol., (1904), pp. 201–292.
[30] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM Rev., 38 (1996), pp. 49–95.
[31] J.A. Walters, *Nonnegative matrix equations having positive solutions*, Math. Comp., (1969), p. 827.